

TOBACCO CONTROL

AN INTERNATIONAL JOURNAL

Editorials

Practical implications of recall bias

Ward and Sanson-Fisher's elegantly simple study of the inaccuracy of patient recall of the tobacco content of a medical encounter is both important and fascinating.¹ It is important because the perception and memory problem it identifies has many practical implications for clinicians, researchers, and all those involved in trying to improve medical care delivery. It is fascinating because, contrary to the expectations of nearly everyone I have talked to, the inaccuracy involved major over-reporting of a medical event, not under-reporting.

However, before considering the implications, let us be more clear about how serious this recall inaccuracy was that they found. Ward and Sanson-Fisher report sensitivity and specificity rates that are misleadingly reassuring—93% and 79% for asking about tobacco use and 92% and 82% for quit advice, respectively. These rates would be felt by most clinicians to be quite acceptable—in the same general range as those for Pap smears or mammography. However, positive predictive value calculations demonstrate that only 56% of the patient reports of being asked about smoking and only 42% of the reports of quit advice were true, for false positive rates of 44% and 58%. Thus, although a negative report of either asking or quit advice is quite valid (negative predictive values of 97% and 99%), a positive report is no better than 50:50. In other words, a patient report of advice to quit smoking during a visit is as likely to be false as true.

What does this study mean for us clinicians? First, it suggests that our smoking discussions or advice may be even less common than has been assumed from the studies of Anda *et al*² or Frank *et al*,³ which showed that only 40–50% of smokers report ever having received advice to quit. When I describe these data to physician audiences, they typically express disbelief, certain that smokers were under-reporting for a variety of reasons. If those surveys also had over-reports of 50%, the true rates might have been much closer to the infrequency with which the Australian trainees apparently asked about smoking (23% of the time) or advised smokers to quit (12%) (as calculated from data in tables 1 and 2 of the paper by Ward and Sanson-Fisher¹). Because the meta-analysis by Kottke *et al*⁴ and many other studies since then have suggested that quit success is directly related to the frequency of advice, we have to stop our false reassurance about what we are doing.

Those who only hear about their smoking this often (once in every three to five encounters) cannot be blamed for feeling that we don't regard it as a very serious problem.

One logical consequence of this infrequency is to make an even stronger case for developing clinic systems that assure consistent identification of tobacco use and consist-

ent reminding to us clinicians that we should try harder to include tobacco use in our encounters with users.⁵ The report of the American Cancer Society's Advisory Group is one of the best summaries of the need for such systems.⁶

This study should also force us to rethink our clinical and research views of the communication problems between patients and clinicians. This topic has been getting a great deal of attention lately with many continuing medical education and research articles focusing on how to improve both communications and the relationship with patients. However, most of those articles focus primarily on the issues of listening more to patients or appreciating that patients only hear or remember a small part of what is really said. This report should remind us all that there is also a very big problem of false perceptions and recollections. Moreover, physicians are human too, and we probably have the same problems.

But are these reporting/recall problems unique to smoking cessation? It seems very unlikely. One piece of evidence is the finding in this study that the very patients who one might predict would be most likely to recall or report incorrectly (smokers) are nearly twice as accurate as non-smokers. Other evidence comes from the large number of studies suggesting that both patient and physician reports are inaccurate for other preventive services as well as from managing clinical problems.

Rohrbaugh and Rogers, for example, conducted an interesting comparison of immediate post-encounter questionnaires completed by five physicians and 189 patients.⁷ Even at a very gross level of reporting on whether specific types of events occurred (prescribed drugs, provided counselling, performed a procedure), there was surprisingly low agreement. The range of agreement was from 57% to 91% with the lowest agreement being for counseling. Boyer *et al* compared patient interview information with what was in the medical record with regard to 121 patients being seen regularly for spondyloarthropathy.⁸ They found that there was substantial under-documentation in the record of information necessary for the diagnosis and equally large inaccuracies in patient reports of such things as the age at onset of back pain. Only 10% of the patients with inflammatory back pain had information in the record allowing a diagnosis to be made, despite the fact that patients had most of that information.

What about other preventive services? Hiatt *et al* compared interview reports of the occurrence of six objective screening tests to what could be verified in their health maintenance organisation (HMO) medical records.⁹

Only for sigmoidoscopy was there greater than 79% concordance between reports and documentation. In other words, there was 25–50% over-reporting by patients of

tests having been done, with the exception of the one screening test that is extremely memorable. Montano and Phillips compared physician self-report with patient survey and chart audits for seven cancer screening tests.¹⁰ They found fair correlation between the chart audit and patient survey (0.53–0.90), but much weaker correlation between physician reports and either of the other two measures (0.14–0.44)

All of these discrepancies pose serious problems for clinicians who are relying on either the patient or the medical record to learn whether patients have a condition or whether they need preventive services. Inaccuracies pose even larger problems for researchers who are attempting to measure whether and when medical events occurred. A saving grace for the researchers may be found in the evidence from both this study and others that there is relatively little difference among patients. Even socioeconomic status does not seem to make very much difference in the rate of inaccurate reporting. Thus, although the actual rates may be inaccurate, it may still be worth studying comparative rates among different practices and as a way to document change over time. For example, it is probably still worthwhile to use patient report of medical behaviour about smoking to learn whether an intervention has made any difference. Ward and Sanson-Fisher's study suggests, however, that if we care about accurate quantification, we should pay more attention to the rate of negative reports than to the rate of positive reports.¹ The negative reports appear to be much more accurate.

Beyond the clinical and research implications of these findings, what meaning does all this have for the burgeoning reliance on data to improve the quality of care? The growing use of continuous quality improvement methods in medical practice places great emphasis on data to establish a baseline, to understand the nature of problems in care, and to measure improvements over time. The findings of Ward and Sanson-Fisher force us to be more cautious about the interpretation of the actual numbers and to rely even more on collecting data from patients. After all, our own estimates of our behaviour are clearly even more inaccurate. We must use patient reports to learn what they want and need as well as what they believe has happened during clinical encounters. In addition, we should be more interested in the change in numbers over time than in their precision if we really want to improve what we do. Thus, we need to rely even more on run charts and control charts as well as on trends in their patterns.

Finally, these findings have implications for the increasing efforts to compare Health Maintenance Organizations or medical care systems on the quality of their care. The first major effort to produce national comparison data about managed care organisations has been introduced during the past few years by the National Committee for Quality Assurance (NCQA). This private accreditation body has developed standardised measures of health plan

performance in a "report card" called HEDIS (Health Plan Employer Data and Information Set). The current version (HEDIS 2.5) has emphasised preventive services by including measures on mammography, Pap and cholesterol tests, and immunisation; and it is very likely that version 3.0, now published in draft form for public comment,¹¹ will contain a measure of the most important clinical preventive service—smoking cessation efforts. Because the most important measure of effectiveness of cessation action is probably the frequency with which tobacco users receive quitting support, and because the best way to obtain this information has been felt to be patient report, NCQA has proposed, as a HEDIS 3.0 measure, the proportion of adult smokers or recent quitters who received advice to quit smoking from a health plan provider during the past year.¹¹ The study by Ward and Sanson-Fisher raises serious questions about this approach, although one could raise equally difficult questions about virtually every other measure in HEDIS.

I hope that this important study does not lead to discarding of this proposed measure. It is so important to emphasise clinical tobacco interventions, and the data from Ward and Sanson-Fisher do not necessarily detract from either the comparability of rates among organisations or the ability to assess change over time. However, it is clear that additional research is needed to better understand the potential for confounding in those comparisons.

LEIF I SOLBERG

Group Health Foundation/HealthPartners,
8100 34 Avenue South, PO Box 1309,
Minneapolis,
Minnesota 55440, USA
email: Leif.I.Solberg@healthpartners.com

- 1 Ward J, Sanson-Fisher R. Accuracy of patient recall of opportunistic smoking cessation advice in general practice. *Tobacco Control* 1996;5:110–113
- 2 Anda RF, Remington PL, Sienko DG, Davis RM. Are physicians advising smokers to quit? The patient's perspective. *JAMA* 1987;257:1916–9.
- 3 Frank E, Winkleby MA, Altman DG, Rockhill B, Fortmann SP. Predictors of physicians' smoking cessation advice. *JAMA* 1991;266:3139–44.
- 4 Kotke TE, Battista R, DeFries G, Breek M. Attributes of successful smoking cessation interventions in medical practice. A meta-analysis of 39 controlled trials. *JAMA* 1988;259:2882–9.
- 5 Fiore MC, Bailey WC, Cohen SJ, et al. *Smoking cessation. Clinical Practice Guideline No. 18*. Rockville, Maryland: US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, 1996. (AHCPR Publication No 96-0692.)
- 6 Leininger LS, Finn L, Dickey L, et al. An office system for organizing preventive services. *Arch Fam Med* 1996;5:108–15.
- 7 Rohrbaugh M, Rogers JC. What did the doctor do? When physicians and patients disagree. *Arch Fam Med* 1994;3:125–9.
- 8 Boyer GS, Templin DW, Goring WP, et al. Discrepancies between patient recall and the medical record: potential impact on diagnosis and clinical assessment of chronic disease. *Arch Intern Med* 1995;155:1868–72.
- 9 Hiatt RA, Perez-Stable EJ, Quesenberry C Jr, Sabogal F, Otero-Sabogal R, McPhee SJ. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Prev Med* 1995;24:278–85.
- 10 Montano DE, Phillips WR. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *Am J Public Health* 1995;85:795–800.
- 11 HEDIS 3.0 draft for public comment. Washington, DC: National Committee for Quality Assurance, July 1996.